Allen Brain Observatory:
Visual Behavior Neuropixels Dataset Cheatsheet

Loading Dataset Summary Tables from Cache

```
cache_dir = path to either the cache on your hard drive or AWS (if using Jupyterhub)
#from hard drive
cache = VisualBehaviorNeuropixelsProjectCache.from_local_cache(cache_dir=cache_dir)
#from AWS
cache = VisualBehaviorNeuropixelsProjectCache.from_local_cache(cache_dir=cache_dir,
use_static_cache=True)
ecephys_sessions_table = cache.get_ecephys_session_table()
behavior_sessions_table = cache.get_behavior_session_table()
```

Summary Tables

table/pandas dataframe

The cache contains 2 tables which provide a summary level view of all the currently released datasets: ecephys_sessions_table, and behavior_sessions_table . These tables contain roughly the same information but are organized differently. Each table is indexed by its corresponding id type

column name	data type	description	ecephys_sessions_table	behavior_sessions_table
abnormal_activity	nan or list	List of experiment time stamps when possible epileptic activity was noted.	x	
abnormal_histology	nan or list	List of brain areas where possible damage was noted in post-hoc imaging.	x	
age_in_days	int	age of mouse in days	x	x
behavior_session_id	int	unique identifier for a behavior session	x	x
channel_count	float	total number of channels on all probes used for experiment	x	
date_of_acquisition	date time object	date and time of data acquisition, yyyy-mm-dd hh:mm:ss.	x	x
ecephys_session_id	int	unique identifier for an ecephys recording session	x	x
equipment_name	string	identifier for equipment data was collected on	x	x
experience_level	string	'Familiar': image set mouse was trained on, 'Novel': not the image set the mouse was trained on.	x	
file_id	int	lookup id to retrieve NWB file from S3 or the local cache.	x	
genotype	string	full genotype of transgenic mouse	x	x
image_set		image set shown for a particular behavior session or ophys session	x	x
mouse_id	int	unique identifier for a mouse	x	x
prior_exposures_to_image_set	float 64	number of prior sessions (during training or ophys) where the mouse was exposed to the image set shown in the current session. Starts at 0 for first exposure	x	x
prior_exposures_to_omissions	int 64	number of sessions where the mouse was exposed to omissions. Starts at 0 for first exposure. Omissions do not occur during training	x	x
prior_exposures_to_session_type	int 64	Number of previous sessions (during training or ophys) where the mouse was exposed to the current session type. Starts at 0 for first exposure	x	
project_code	string	'NeuropixelVisualBehavior': Project this session belongs to	x	
session_number	float 64	[1, 2] Indicates whether this session was the first or second recording day for the mouse. Takes values '1' or '2'.	x	x
session_type	string	Visual stimulus type displayed during behavior session	x	x
sex	string	['M', 'F'] Sex of the mouse	x	x
structure_acronyms	string	List of CCF structures recorded during this experiment	x	
unit_count	float	number of units recorded during this	x	

Loading Session Objects

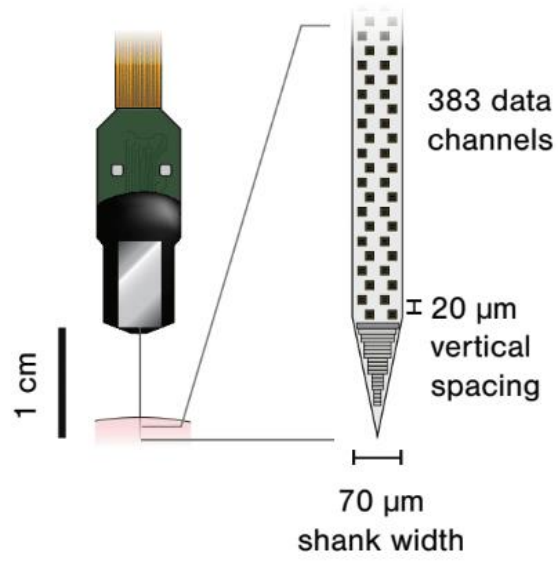
```
ecephys_session_id = ecephys_sessions_table.index.values[i]
session = cache.get_ecephys_session(ecephys_session_id= ecephys_session_id)
```

Session Attributes

color coded by data type

attributes and methods associated with an ecephys session dataset object.

eye tracking	table	ellipse fit parameters for eye tracking
licks	table	a dataframe containing lick times
mean_waveforms	dictionary	Dictionary containing the mean waveform across all probe channels for every unit.
metadata	dictionary	dataset specific metadata including genotype, imaging depth, etc.
optotagging_table	table	Dataframe containing laser waveform parameters and onset times for trials.
probes	table	Dataframe containing metadata for the probes inserted during this session.
running speed	table	a table containing timestamps and filtered running speed
raw running speed	table	a table containing timestamps and unfiltered running speed
rewards	table	timestamps and delivered rewards
reward_rate	array	Calculated reward rates
spike_amplitudes	dictionary	spike amplitudes (in Volts) for every unit recorded during this session
spike_times	dictionary	spike times for every unit recorded during this session
stimulus presentations	table	times of every stimulus presentation/image flash as well as associate metadata
stimulus templates	table	a table containing stimulus images presented
stimulus timestamps	array	timestamps associated with the stimulus presentations on the monitor
task parameters	dictionary	parameters used to define task runtime behavior
trials	table	behavior trial data including change time and trial type
Units	Table	Dataframe containing quality and waveform metrics for every unit recorded during the session
name	string	please ignore



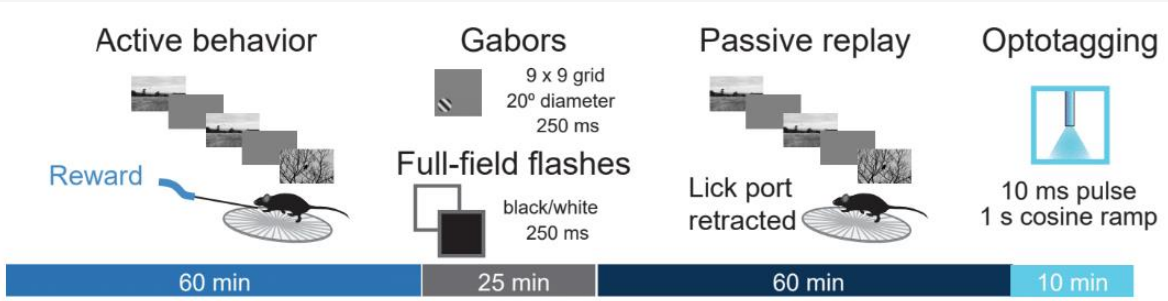
Available Sessions

Experience Level	Genotype		
	SST	VIP	WT
Familiar	31	22	25
Novel	28	23	24

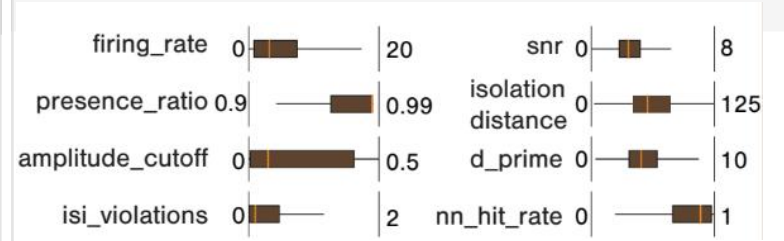
Available Units (no qc filters applied)

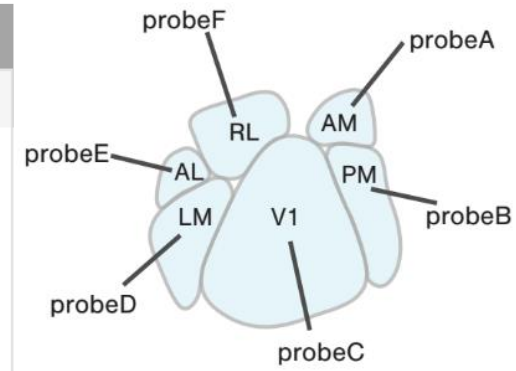
ANATOMICAL STRUCTURE	Full Name	Acronym	Total Units
VISUAL CORTEX	Primary visual cortex	VISp	21567
	lateromedial area	VISl	19252
	rosterolateral area	VISrl	15319
	Anterolateral area	VISal	17442
	Posteromedial area	VISpm	19772
	Aneromedial area	VISam	16938
HIPPOCAMPAL FORMATION	Cornu ammonis 1	CA1	42687
	Cornu ammonis 3	CA3	10834
	Denate gyrus	DG	22484
	Subiculum	SUB	8319
	Posubiculum	ProS	4859
THALAMUS	Lateral geniculate nuc.	LGd	3866
	Lateral posterior nuc.	LP	5479
MIDBRAIN	Midbrain reticular nuc.	MRN	7880

Session Overview



Unit Quality Metrics





session.metadata — dictionary

session-specific metadata including genotype, equipment name etc

key	Value dtype	value
age_in_days	int	age of mouse in days
behavior_session_id	int	unique identifier for a behavior session
behavior_session_uuid	uuid	unique identifier for behavior session
cre_line	string	cre driver line for a transgenic mouse
date_of_acquisition	datetime object	date and time of experiment acquisition, yyyy-mm-dd hh:mm:ss.
driver_line	list of strings	all driver lines for transgenic mouse
ecephys_session_id	int	unique identifier for ephys session
equipment_name	string	identifier for equipment data was collected on
full_genotype	string	full genotype of transgenic mouse
mouse_id	int	unique identifier for mouse
reporter_line	string	reporter line for transgenic mouse
session_type	string	visual stimulus type displayed during behavior session
sex	string	sex of the mouse ['M', 'F']
stimulus_frame_rate	float	rate (in Hz) visual stimulus was displayed

session.probes — table/pandas dataframe

Dataframe containing metadata for the probes inserted during this session.

column name	data type	description
id [INDEX]	int	Unique identifier for probe insertion
name	string	Name of probe indicating which of the six probe positions this probe filled (A-F).
sampling_rate	float	Sampling rate for AP band data
has_lfp_data	bool	Flag indicating whether LFP data was collected on this probe; ignore for now
lfp_sampling_rate	float	Sampling rate for LFP data; ignore for now
location	string	Locations are defined at the channel level. Ignore.

session.optotagging_table — table/pandas dataframe

Dataframe containing laser waveform parameters and onset times for every trial during the optotagging stimulus.

column name	data type	description
id [INDEX]	int	Trial index
condition	string	Description of laser waveform used for this trial
duration	float	Duration of laser stimulation for this trial
level	float	Laser command voltage for this trial (higher values indicate higher power laser stimulation)
start_time	float	Start time of laser stimulation for this trial
stimulus_name	string	Name of laser waveform used for this trial
stop_time	float	Time when laser stimulation stops for this trial

session.mean_waveforms — dictionary

Key	Value dtype	values
units id (index values of the units table)	array of floats	2D waveform (384 channels, 84 samples) for each unit

session.get_units() — table/pandas dataframe

Returns units dataframe for this session. This dataframe stores quality and waveform metrics for every unit recorded during this session.

column name	data type	description
id [INDEX]	int	Unique identifier for each unit. These ids are the keys to the mean_waveforms, spike_times and spike amplitudes tables
amplitude	float	Peak to trough amplitude for mean 1D waveform in microvolts
amplitude_cutoff	float	estimate of miss rate based on amplitude histogram (ie fraction of spikes estimated to have been below detection threshold)
cluster_id	int	Unique identifier for cluster to which the spike for this unit were assigned by Kilosort 2
cumulative_drift	float	cumulative change in spike depth along probe throughout the recording
d_prime	float	classification accuracy based on LDA
firing_rate	float	Mean firing rate over entire recording
isi_violations	float	Ratio of refractory violation rate to total spike rate
isolation_distance	float	Distance to nearest cluster in Mahalanobis space
l_ratio	float	The Mahalanobis distance and chi-squared inverse cdf are used to find the probability of cluster membership for each spike
max_drift	float	Maximum change in unit depth across recording
nn_hit_rate	float	Fraction of nearest neighbors in PCA space for spikes in unit cluster that are also in unit cluster
peak_channel_id	int	Unique channel identifier for channel on which the given unit's waveform was largest amplitude. This id is the index of the channels table.
presence_ratio	float	Fraction of time during session for which a unit was spiking
PT_ratio	float	Ratio of the max (peak) to the min (trough) amplitudes for 1D waveform
quality	string	['good', 'noise'] Label assigned based on waveform shape as described here. Either 'good' for physiological waveforms or 'noise' for artifactual waveforms.
recovery_slope	float	Slope of recovery of 1D waveform to baseline after repolarization (coming down from peak)
repolarization_slope	float	Slope of repolarization of 1D waveform to baseline after trough
silhouette_score	float	Standard metric for cluster overlap, computed in PCA space
snr	float	signal-to-noise ratio for 1D waveform
spread	float	Distance along probe shank in microns for which the spike amplitude was above 12% of the peak channel amplitude
velocity_above	float	Slope of spike propagation velocity traveling in dorsal direction from soma (note to avoid infinite values, this is actaully the inverse of velocity: ms/mm)
velocity_below	float	Slope of spike propagation velocity traveling in ventral direction from soma (note to avoid infinite values, this is actually the inverse of velocity: ms/mm)
waveform_duration	float	Time from trough to peak for 1D waveform in milliseconds
local_index	int	This column can be ignored.

session.spike_amplitudes — dictionary

spike amplitudes (in Volts) for every unit recorded during this session.

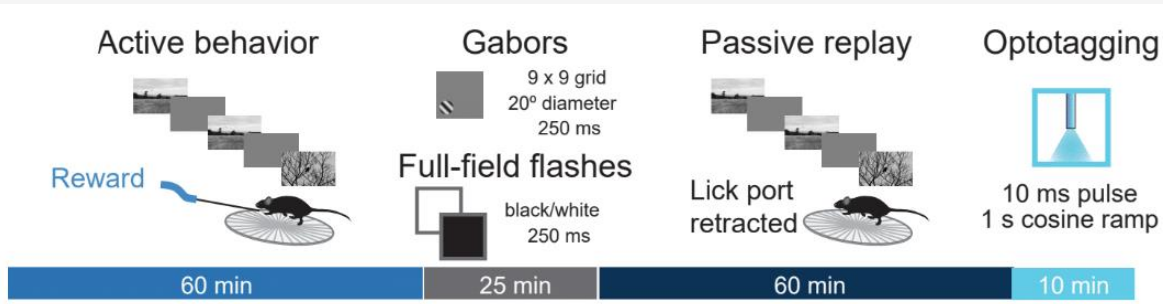
key	Value dtype	values
units id (index values of the units table)	array of floats	amplitudes in volts

session.spike_times — dictionary

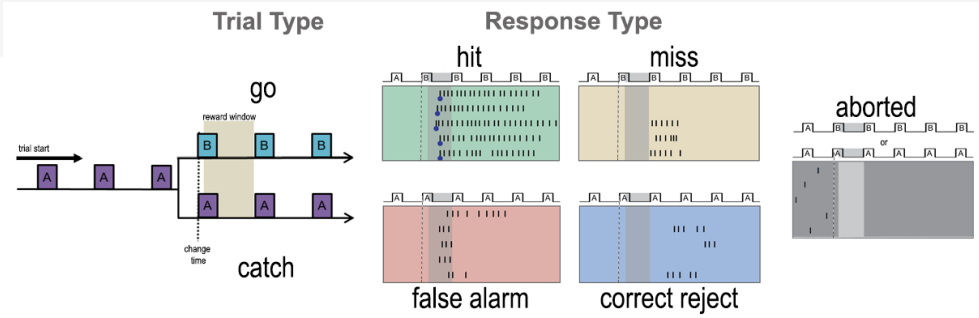
spike times for every unit recorded during this session

key	Value dtype	values
units id (index values of the units table)	array of floats	spike times in seconds

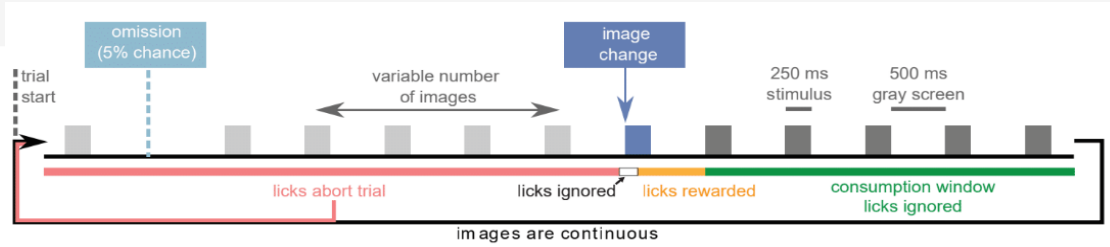
Session Overview



Trial Structure



Change Detection Task Structure



session.stimulus_presentations

table/pandas dataframe

A dataframe containing the times of every visual stimulus presentation as well as associated metadata

column	data type	description
stimulus_presentations_id [INDEX]	int	identifier for a stimulus presentation (presentation of an image)
active	bool	Boolean indicating the stimulus block during which the mouse was performing the task
color	float	Color of full field flash (-1 indicates black, 1 indicates white). Valid for stimulus block 4.
contrast	float	Contrast of full field flash. Valid for stimulus block 4.
duration	float	duration of image flash in seconds (stop time - start time), NaN if omitted
end_frame	int	image presentation end frame
flashes_since_change	float	Number of image flashes since the last image change. Valid for stimulus blocks 0 and 5.
image_name	string	name of image presented, if 'omitted' then no image was presented. Valid for blocks 0 and 5.
is_change	bool	True if the image changed identity on this flash. Valid for blocks 0 and 5.
omitted	bool	True if no image was shown for this stimulus presentation. Valid for blocks 0 and 5.
orientation	float	Orientation of gabor stimulus. Valid for block 2.
position_x	float	Position of gabor stimulus in azimuth relative to center of screen in degrees. Valid for block 2.
position_y	float	Position of gabor stimulus in elevation relative to center of screen in degrees. Valid for block 2.
rewarded	bool	True if the mouse was rewarded after this image flash. Valid for blocks 0 and 5. For block 5, True if the mouse was rewarded for the corresponding flash during block 0.
spatial_frequency	float	Spatial frequency of gabor in cycles/degree. Valid for block 2.
start_frame	int	image presentation start frame
start_time	float	image presentation start time in seconds
stimulus_block	int	Index of stimulus block for this stimulus. 0 is active behavior, 1 is spontaneous activity, 2 is gabor stimuli, 3 is spontaneous activity, 4 are full-field flashses and 5 is passive replay.
stimulus_index	float	Index of receptive field mapping stimulus used. Either 1 for gabors or 2 for full-field-flashes. NaN for all other stimulus blocks.
stimulus_name	string	Name of stimulus group for this stimulus
stop_time	float	image presentation end time in seconds
temporal_frequency	float	Temporal frequency in Hz of gabor stimulus. Valid for block 2.

session.trials

table/pandas dataframe

behavior trial data including change time and trial type

column name	data type	description
trials_id [INDEX]	Int	The trial number/index
aborted	bool	[behavior response type] True if mouse licks before the scheduled change time
auto_rewarded	bool	True if free reward was delivered for that trial. Occurs during the first 5 trials of a session and throughout as needed.
catch	bool	[trial type] True if there was not a change in stimulus image identity on this trial
change_frame	float	frame of image change
change_image_name	string	name of image that is changed to at the change time, on go trials
change_time_no_display_delay	float	Time of the change stimulus according to the task control computer. This does not account for the time taken for the image to appear on the monitor.
correct_reject	bool	[behavior response type] on a catch trial, mouse either does not lick at all or licks after reward window
false_alarm	bool	[behavior response type] on a catch trial, mouse licks within reward window
go	bool	[trial type] True if there was a change in stimulus image identity on this trial
hit	bool	[behavior response type] on a go trial, mouse licks within reward window
initial_image_name	string	name of image presented at start of trial
lick_times	array of float	array of lick times in seconds during that trial. Empty array if no licks occurred during that trial. Note that these are the lick times as registered by the task control computer. For more accurate lick times, use the licks dataframe.
miss	bool	[behavior response type] on a go trial, mouse either does not lick at all, or licks after reward window
response_time	float	Experiment time of the first lick recorded by the task control computer for each trial.
reward_time	NaN or float	Time the reward is delivered following a correct response or on autorewarded trials.
reward_volume	float	volume of reward in ml. (0.005 for auto reward and 0.007 for earned reward)
start_time	float	start time of the trial in seconds
stimulus_change	bool	True if an image change occurs during the trial (if the trial was both a 'go' trial and the trial was not aborted)
stop_time	float	end time of the trial in seconds
trial_length	float	duration of trial, in seconds (stop time - start time)

Stimulus Block Diagram

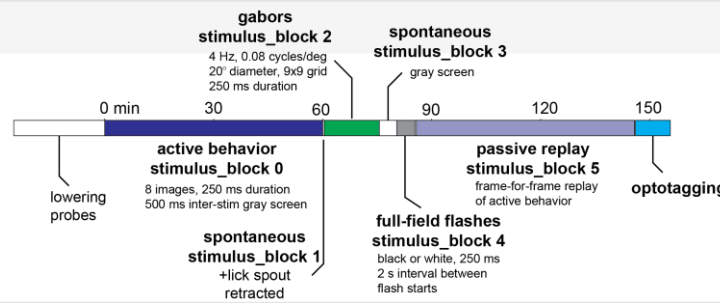
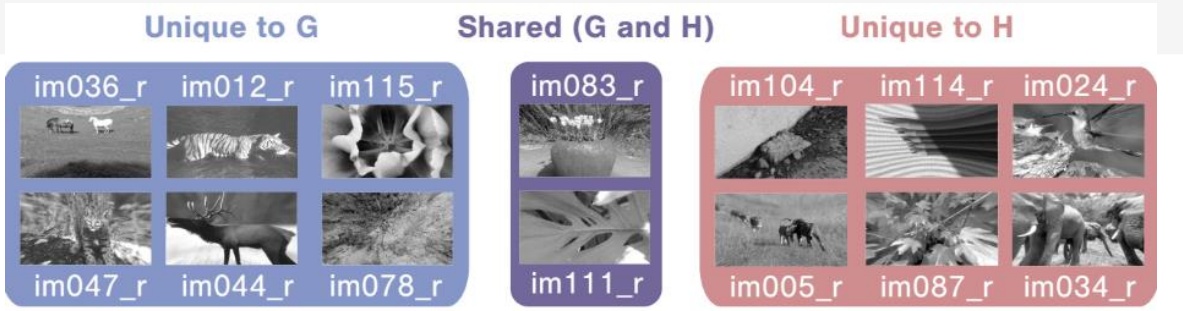


Image Sets



session.task_parameters — dictionary

parameters used to define the task runtime behavior

key	Value dtype	value
auto_reward_volume	float	volume of auto rewards in ml (0.005ml)
blank_duration_sec	list of float	duration in seconds of inter stimulus interval
n_stimulus_frames	int	total number of visual stimulus frames presented during a behavior session
omitted_flash_fraction	float	fraction of stimulus image presentations that are omitted
response_window_sec	list of float	the range of the period following an image change, in seconds, where mouse response influences trial outcome. The first value represents response window start. The second value represents response window end. The values represent time before display lag is accounted for and applied.
reward_volume	float	volume of individual water reward in ml. 0.007 if earned reward, 0.005 if auto reward
session_type	string	visual stimulus type displayed during behavior session
stimulus	string	stimulus type
stimulus_distribution	string	distribution for drawing change times. Either ‘exponential’ or ‘geometric’.
stimulus_duration_sec	float	duration in seconds of each stimulus image presentation
task	string	type of visual stimulus task, typically 'change detection'

session.stimulus_templates — table/pandas dataframe

Dataframe with columns arrays of both the warped and unwarped images

column name	data type	description
image_name [INDEX]	string	Name of natural image
unwarped	array of int	image array of unwarped stimulus image
warped	array of int	image array of warped stimulus image

session.stimulus_timestamps — numpy array

Timestamps associated with the stimulus presentation on the monitor. Sampled at 60Hz. An array of float 64.

session.rewards — table/pandas dataframe

A dataframe containing timestamps of delivered rewards in absolute sync time. Timestamps are sampled at 60 Hz.

column name	data type	description
timestamps	float 64	time in seconds
volume	float 64	volume of individual water reward in ml.
auto_rewarded	bool	True if free reward was delivered for that trial. Occurs during the first 5 trials of a session.

session.get_reward_rate() — numpy array

array of calculated reward rates. Reward rate is calculated of a 25 trial rolling window and provides a measure of the rewards earned per unit time (in unites of rewards/min). Our standard is to define 'engaged' behavior as trials with a reward rate > 2.

session.licks — table/pandas dataframe

A dataframe containing lick timestamps and frames, sampled at 60Hz.

column name	data type	description
timestamps	float 64	time of a lick, in seconds
frame	int	frame of lick

session.running_speed — table/pandas dataframe

Running speed and timestamps sampled at 60hz. A 10Hz low pass filter has been applied to the data. To get the running speed without the filter, use `raw_running_speed`.

column name	data type	description
speed	float 64	speed in cm/sec
timestamps	float 64	time in seconds

session.raw_running_speed — table/pandas dataframe

Unfiltered running speed and timestamps sampled at 60Hz.

column name	data type	description
speed	float 64	speed in cm/sec
timestamps	float 64	time in seconds

session.eye_tracking — table/pandas dataframe

A dataframe containing ellipse fit parameters for the eye, pupil and corneal reflection (cr). Fits are derived from tracking points from a DeepLabCut model applied to video (collected at 30hz) frames of a subject’s right eye. Raw tracking points and raw video frames are not publicly available.

column name	data type	description
frame [index]	int 64	frame of eye tracking video
timestamps	float 64	time in seconds
likely_blink	float 64	True if frame has outlier ellipse fits, which is often caused by blinking / squinting of the eye.
eye_area	float 64	area of eye ellipse in pixels squared
eye_area_raw	float 64	pupil area with no outliers/likely blinks removed.
eye_center_x	float 64	center of eye ellipse on x axis in pixels
eye_center_y	float 64	center of eye ellipse on y axis in pixels
eye_height	float 64	eye height (minor axis of the eye ellipse) in pixels
eye_width	float 64	eye width (major axis of the eye ellipse) in pixels
eye_phi	float 64	eye rotation angle in radians
pupil_area	float 64	area of pupil ellipse in pixels squared
pupil_area_raw	float 64	pupil area with no outliers/likely blinks removed.
pupil_center_x	float 64	center of pupil ellipse on x axis in pixels
pupil_center_y	float 64	center of pupil ellipse on y axis in pixels
pupil_height	float 64	pupil height (minor axis of the pupil ellipse) in pixels
pupil_width	float 64	pupil width (major axis of the pupil ellipse) in pixels
pupil_phi	float 64	pupil rotation angle in radians
cr_area	float 64	area of corneal reflection ellipse in pixels squared
cr_area_raw	float 64	corneal reflection area with no outliers/likely blinks removed.
cr_center_x	float 64	center of corneal reflection on x axis in pixels
cr_center_y	float 64	center of corneal reflection on y axis in pixels
cr_height	float 64	corneal reflection height (minor axis of the CR ellipse) in pixels
cr_width	float 64	corneal reflection width (major axis of the CR ellipse) in pixels
cr_phi	float 64	corneal reflection rotation angle in radians